Exploring NK Fitness Landscapes Using Imitative Learning

José F. Fontanari

Instituto de Física de São Carlos, Universidade de São Paulo,
Caixa Postal 369, 13560-970 São Carlos, São Paulo, Brazil

The idea that a group of cooperating agents can solve problems more efficiently than when those agents work independently is hardly controversial, despite our obliviousness of the conditions that make cooperation a successful problem solving strategy. Here we investigate the performance of a group of agents in locating the global maxima of NK fitness landscapes with varying degrees of ruggedness. Cooperation is taken into account through imitative learning and the broadcasting of messages informing on the fitness of each agent. We find a trade-off between the group size and the frequency of imitation: for rugged landscapes, too much imitation or too large a group yield a performance poorer than that of independent agents. By decreasing the diversity of the group, imitative learning may lead to duplication of work and hence to a decrease of its effective size. However, when the parameters are set to optimal values the cooperative group substantially outperforms the independent agents.

I. INTRODUCTION

The benefits of cooperative work were explored by nature well before the advent of the human species as attested by the collective structures built by slime molds and social insects [1]. In the human context, the socio-cognitive niche hypothesis purports that hominin evolution relies so heavily on social elements to the point of viewing a band of hunter–gatherers as a 'group-level predator' [2]. Whether physical or social, those structures and organizations may be thought of as the organisms' solutions to the problems that endanger their existence (see, e.g., [3]) and have motivated the introduction of the concept of social intelligence in the scientific milieu [4]. In the computer science circle, that concept prompted the proposal of optimization heuristics based on social interaction, such as the popular particle swarm optimization algorithm [5, 6] and the perhaps lesser-known adaptive culture heuristic [7, 8].

Despite the prevalence of the notion that cooperation can aid a group of agents to solve problems more efficiently than if those agents worked in isolation, and of the success of the social interaction heuristics in producing near optimal solutions to a variety of combinatorial optimization problems, we know little about the conditions that make cooperative work more efficient than independent work. In particular, we note that since cooperation (or communication, in general) reduces the diversity or heterogeneity of the group of agents, it may actually end up reducing the efficiency of group work [9, 10]. Efficiency here means that the time to solve a problem scales superlinearly with the number of individuals or resources employed in the task.

In this contribution we study the performance of a group of cooperative agents following a first-principle research strategy to study cooperative problem solving set forth by Huberman in the 1990s that consists of tackling easy tasks, endowing the agents with simple random trial-and-test search tools, and using plain protocols of collaboration [11, 12]. Here the task is to find the global maxima of three realizations of the NK-fitness landscape characterized by different degrees of ruggedness (i.e., values of the parameter K). We use a group of M agents which, in addition to the ability to perform individual trial-and-test searches, can imitate a model agent – the best performing agent at the trial – with a probability p. Hence our model exhibits the two critical ingredients of a collective brain according to Bloom: imitative learning and a dynamic hierarchy among the agents [3]. The model exhibits also the key feature of distributed cooperative problem solving systems, namely, the exchange of messages between agents informing each other on their partial success (i.e., their fitness at the current trial) towards the completion of the goal [11].

We find that the presence of local maxima in the fitness landscape introduces a complex trade-off between the computational cost to solve the task and the group size M. In particular, for a fixed frequency of imitation p, there is an optimal value of M at which the computational cost is minimized. This finding leads to the conjecture that the efficacy of imitative learning could be a factor determinant of the group size of social animals [13].

Our study departs from the vast literature on cooperation that followed Robert Axelrod's 1984 seminal book The Evolution of Cooperation [14] since in that game theoretical approach it is usually assumed a priori that mutual cooperation is the most rewarding strategy to the individuals. On the other hand, here we consider a problem solving scenario and a specific cooperation mechanism (imitation) aiming at determining in which conditions cooperation is more efficient than independent work.

The rest of the paper is organized as follows. In Section II we offer a brief overview of the NK model of rugged fitness landscapes. In Section III we present a detailed account of the imitative learning search strategy and in Section IV we study its performance on the task of finding the global maxima of three realizations of the NK landscape with $N = 12$ and ruggedness parameters $K = 0$, $K = 2$ and $K = 4$. The rather small size of the solution

space ($2^{12} = 4096$ binary strings of length 12) allows the full exploration of the space of parameters and, in particular, the study of the regime where the group size is much greater than the solution space size. Finally, Section V is reserved to our concluding remarks.

II. NK MODEL OF RUGGED FITNESS LANDSCAPES

The NK model of rugged fitness landscapes introduced by Kauffman [15, 16] offers the ideal framework to test the potential of imitative learning in solving optimization problems of varied degrees of difficulty, since the ruggedness of the landscape can be tuned by changing the two integer parameters – N and K – of the model. Roughly speaking, the parameter N determines the size of the solution space whereas the value of $K = 0, \ldots, N-1$ influences the number of local maxima and minima on the landscape.

The solution space consists of the 2^N distinct binary strings of length N, $\mathbf{x} = (x_1, x_2, \ldots, x_N)$ with $x_i = 0, 1$. To each string we associate a fitness value $\Phi(\mathbf{x})$ which is an average of the contributions from each component i in the string, i.e.,

$$\Phi(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^{N} \phi_i(\mathbf{x}), \qquad (1)$$

where ϕ_i is the contribution of component i to the fitness of string $\mathbf{x}$. It is assumed that ϕ_i depends on the state x_i as well as on the states of the K right neighbors of i, i.e., $\phi_i = \phi_i(x_i, x_{i+1}, \ldots, x_{i+K})$ with the arithmetic in the subscripts done modulo N. In addition, the functions ϕ_i are N distinct real-valued functions on $\{0, 1\}^{K+1}$. As usual, we assign to each ϕ_i a uniformly distributed random number in the unit interval [15, 16]. Hence $\Phi \in (0, 1)$ has a unique global maximum.

For $K = 0$ there are no local maxima and the sole maximum of Φ is easily located by picking for each component i the state $x_i = 0$ if $\phi_i(0) > \phi_i(1)$ or the state $x_i = 1$, otherwise. For $K > 0$, finding the global maximum of the NK model is a NP-complete problem [17], which essentially means that the time required to solve the problem using any currently known deterministic algorithm increases exponentially fast as the size N of the problem grows [18]. The increase of the parameter K from 0 to $N-1$ decreases the correlation between the fitness of neighboring configurations (i.e., configurations that differ at a single component) in the solution space and for $K = N-1$ those fitness values are uncorrelated so the NK model reduces to the Random Energy model [19, 20]. To illustrate the effect of varying K on the ruggedness of the fitness landscape, in Fig. 1 we show the relative fitness of a string Φ/Φ_g as function of its Hamming distance d to the global maximum for $N = 25$ and different values of K. Here Φ_g stands for the fitness of the global maximum. For each K, the figure shows the
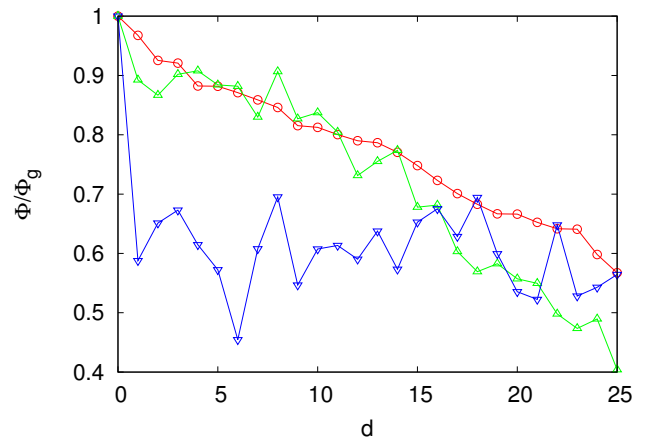

FIG. 1. (Color online) Normalized fitness as function of the Hamming distance to the global maximum for $N = 25$ and $K = 0$ (*circles*), $K = 5$ (*triangles*) and $K = 20$ (*inverted triangles*). The lines are guides to the eye.

results for a single realization of the fitness landscape and for a single trajectory in the solution space that begins at the maximum ($d = 0$) and changes the state components sequentially until all N states are reversed ($d = N$).

We note that the ruggedness of the landscapes (essentially, the number of local maxima) can vary wildly between landscapes characterized by the same values of N and $K > 0$ [15, 16] and so can the performance of any search heuristic based on the local correlations of the fitness landscape. Hence in order to appreciate the role of the parameters that are relevant to imitative learning, namely, the group size M and the imitation probability p, for fixed N and K we consider a single realization of the NK fitness landscape.

III. IMITATIVE LEARNING SEARCH

We assume a group or system composed of M agents. Each agent operates in an initial binary string drawn at random with equal probability for the digits 0 and 1. In the typical situation that the size of the solution space is much greater than the group size $2^N \gg M$ we can consider those initial strings as distinct binary strings, but here we will consider the case that $M \gg 2^N$ as well, so that many copies of a same string are likely to appear in this initial stage of the simulation. In addition, we assume that the agents operate in parallel.

At any trial t, each agent can choose with a certain probability between two distinct processes to operate on the strings. The first process, which happens with probability $1 - p$, is the elementary or minimal move in the solution space, which consists of picking a string bit $i = 1, \ldots, N$ at random with equal probability and then flipping it. The repeated application of this operation is capable of generating all the 2^N binary strings starting

from any arbitrary string. The second process, which happens with probability p, is the imitation of a model string. We choose the model string as the highest fitness string in the group at trial t. The string to be updated (target string) is compared with the model string and the different digits are singled out. Then the agent selects at random one of the distinct bits and flips it so that the target string becomes now more similar to the model string. The parameter $p \in [0, 1]$ is the imitation probability and the case $p = 0$ corresponds to the baseline situation in which the M agents explore the solution space independently.

The specific imitation procedure proposed here was inspired by the mechanism used to model the influence of an external media [21–24] in the celebrated agent-based model proposed by Axelrod to study the process of culture dissemination [25]. It is important to note that in the case the target string is identical to the model string, and this situation is not uncommon since the imitation process reduces the diversity of the group, the agent executes the elementary move with probability one. This procedure is different from that used in [13], in which strings identical to the model string are not updated within the imitation process. However, both procedures yield qualitatively similar results, except in the regime where imitation is extremely frequent, i.e., for $p \approx 1$. In particular, in the present implementation, a small group can find the global maximum in the case $p = 1$ since the model string can explore the solution space through the elementary move whereas the other strings are simply followers.

The collective search ends when one of the agents finds the global maximum and we denote by t^* the number of trials made by the agent that found the solution. Of course, t^* stands also for the number of trials made by any one of the M agents, since they operate in parallel and the search halts simultaneously for all agents. In other words, the trial number t is updated, or more specifically, incremented by one unit, when the M agents have executed one of the two processes on its associated string. We note that except for the case $p = 0$, the update of the M agents is not strictly a parallel process since the model strings may change several times within a given trial. Nonetheless, since in a single trial all agents are updated, the total number of agent updates at trial t is given by the product Mt.

The efficiency of the search strategy is measured by the total number of agent updates necessary to find the global maximum (i.e., Mt^*) which can then be interpreted as the computational cost of the search. In addition, since we expect that the typical number of trials to success t^* scales with the size of the solution space 2^N, we will present the results in terms of the rescaled computational cost, defined as $C \equiv Mt^*/2^N$.

An interesting variant of our imitation rule is obtained by relaxing the condition that only the string with the highest fitness value can be imitated and allowing any string to pose as a model according to an imitation prob-
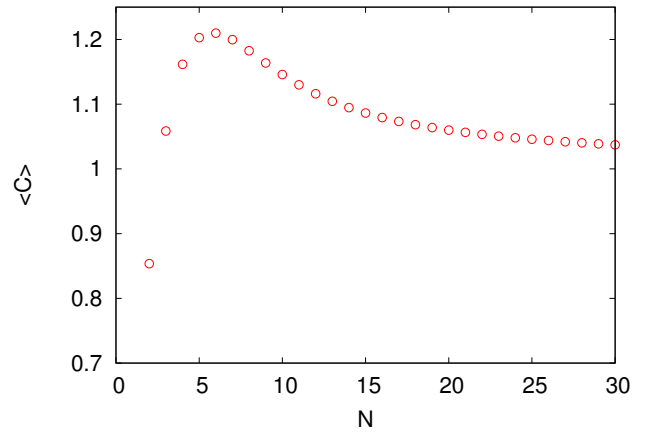

FIG. 2. (Color online) Single-agent mean rescaled computational cost $\langle C \rangle$ as function of the binary string length N that determines the size of the solution space 2^N. These results, which were obtained by the numerical calculation of the eigenvalues of the tridiagonal stochastic matrix $\mathbf{T}$, do not depend on the ruggedness of the landscape.

ability function proportional to the string relative fitness (e.g., a Fermi function). Such fitness dependent imitation probability functions are frequently used in the study of cooperation dilemmas [26, 27] and may be useful to prevent the string population being trapped in the local maxima. In this paper, however, we will consider only the noiseless or zero-temperature limit of those imitation functions, where only the string with the maximum relative fitness can be imitated.

IV. RESULTS

Here we report the results of extensive Monte Carlo simulations of groups of agents that use imitative learning to search for the global maxima of three representative realizations of the NK fitness landscape. For fixed N and K we generate a single realization of the fitness landscape and carry out 10^5 searches in order to determine the dependence of the mean rescaled computational cost $\langle C \rangle$ on the imitation probability p and on the group size M. In addition, for the baseline case in which the M agents explore the landscape independently ($p = 0$) we derive an analytical expression for the computational cost.

A. Independent search

In this case there is no imitation ($p = 0$) and the ruggedness of the landscape has no effect on the performance of the search, which depends only on the size of the solution space, 2^N. Because of this independence on the landscape we can derive exact results for the time

needed for M independent agents to find the global maximum. For this analysis we can assume that the global maximum is the string with the N digit values equal to 1, i.e., $(1, 1, \ldots, 1)$, without loss of generality.

Let us consider first the case of a single agent that operates on a string with j digits 1. According to the elementary move, the probability that the resulting string has i digits 1 is

$$T_{ij} = (1 - j/N)\,\delta_{i,j+1} + j/N\delta_{i,j-1} \qquad (2)$$

for $j = 1, \ldots, N - 1$, $T_{i0} = \delta_{i,1}$, and $T_{iN} = \delta_{i,N}$, where $\delta_{i,j}$ is the Kronecker delta. Since $0 \leq T_{ij} \leq 1$ and $\sum_i T_{ij} = 1$ for $j = 0, \ldots, N$ the matrix $\mathbf{T}$ is a tridiagonal stochastic matrix. The associated stochastic process has a single absorbing state $i = N$ and $\pi_N(t) = \sum_{i=0}^{N} T_{Ni}\pi_i(t-1)$ yields the probability that the agent finds the solution before or at trial t. For large t this quantity is given $\pi_N(t) = 1 - (\lambda_N)^t$ where λ_N is the second largest eigenvalue of $\mathbf{T}$. (The largest eigenvalue is 1 because $\mathbf{T}$ is a stochastic matrix.) The probability that the agent finds the solution exactly at trial $t = t^*$ is $p_N(t^*) = \pi_N(t^*) - \pi_N(t^* - 1)$, i.e.,

$$p_N(t^*) = (\lambda_N)^{t^*-1}(1 - \lambda_N), \qquad (3)$$

which is a geometric distribution with success probability $1 - \lambda_N$. The value of λ_N can be easily obtained numerically (see, e.g., [28]) provided N is not too large since $2^N(1 - \lambda_N)$ is on the order of the unity and so it becomes practically impossible to distinguish λ_N from 1 for $N > 30$. The mean rescaled computational cost $\langle C \rangle = 1/[2^N(1 - \lambda_N)]$ is shown in Fig. 2 for different string lengths N. (Recall that $M = 1$ at this stage.) In particular for $N = 12$ we find $\lambda_{12} \approx 0.99978$ that implies $\langle C \rangle \approx 1.11612$. For very large N, the probability of success becomes $1 - \lambda_\infty = 1/2^N$. We note that although eq. (3) is valid strictly for large t^* only, the fact that t^* is typically on the order of 2^N makes this geometric distribution an exceedingly good approximation to the correct distribution of absorbing times.

Now let us consider the case of $M = 2$ agents searching independently for the global maximum. Since the process halts when one of the agents finds the global maximum, the halting time is $t_2^* = \min\{t_a^*, t_b^*\}$ where t_a^* and t_b^* are independent random variables distributed by the geometric distribution (3). It is easy to show that the distribution of t_2^* is also geometric with probability of success $1 - (\lambda_N)^2$ [29]. In the case of $M = 3$ agents the halting time $t_3^* = \min\{t_2^*, t_a^*\}$ and since both t_2^* and t_a^* are geometrically distributed, though with distinct success probabilities, we find that t_3^* is also geometrically distributed with probability of success $1 - (\lambda_N)^3$. The generalization of this reasoning to M agents yields that the mean scaled cost is

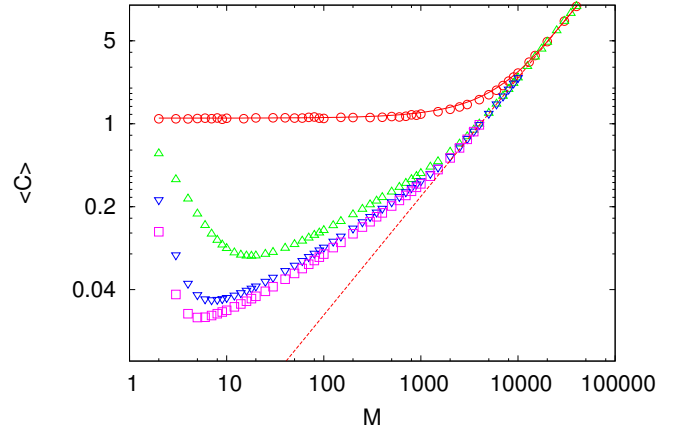$$\langle C \rangle = \frac{M}{2^N\left[1 - (\lambda_N)^M\right]}. \qquad (4)$$

FIG. 3. (Color online) Mean rescaled computational cost $\langle C \rangle$ as function of the group size M for the imitation probability $p = 0$ (*circles*), $p = 0.5$ (*triangles*), $p = 0.8$ (*inverted triangles*), and $p = 1$ (*squares*). The solid curve is eq. (4) and the dashed line is the linear function $\langle C \rangle = M/2^N$. The parameters of the NK landscape are $N = 12$ and $K = 0$. The landscape exhibits a single maximum.

Since λ_N is close to 1 we can write $(\lambda_N)^M \approx e^{-M(1-\lambda_N)}$ so that $\langle C \rangle \approx 1/\left[2^N(1 - \lambda_N)\right]$ for $M \ll 1/(1 - \lambda_N)$ and $\langle C \rangle \approx M/2^N$ for $M \gg 1/(1 - \lambda_N)$. We recall that for $N = 12$ we have $1/(1 - \lambda_{12}) \approx 4545$. As expected, eq. (4) matches the simulation data very well (see Fig. 3).

B. Search on a Smooth Landscape ($K = 0$)

The NK fitness landscape with $K = 0$ exhibits a single global maximum and no local maxima. The results of the performance of the imitative search for a landscape with $N = 12$ and $K = 0$ are summarized in Fig. 3.

As shown in the previous subsection, the mean computational cost for non-interacting agents ($p = 0$) is a constant $\langle C \rangle \approx 1.12$ provided the group size M is not too large compared to $1/(1 - \lambda_{12}) \approx 4545$, which is close to the size of the solution space, $2^{12} = 4096$. In the case the group size M is very large, the agents begin to duplicate each other's work leading to a linear increase of $\langle C \rangle$ with increasing M. More pointedly, in this regime we find $\langle C \rangle = M/2^N$ (see the straight line in Fig. 3). We stress that a constant computational cost means that the time t^* the group requires to find the global maximum decreases with the inverse of the group size (i.e., the time to solve the problem decreases linearly with the number of agents), whereas a computational cost that grows linearly with the group size means that t^* does not change as more agents are added to the group.

Allowing the agents to imitate the best performer (model) at each trial leads to a rapid reduction of the computational cost provided the group size remains small. The best performance is achieved for $M = 5$ and $p = 1$ and corresponds to a fiftyfold decrease of the computational

TABLE I. Local maxima for the studied realization of the NK fitness landscape with $N = 12$ and $K = 2$

Maximum	Φ/Φ_g	d
1	0.7681	5
2	0.9114	5
3	0.9204	3
4	0.9801	3

cost with respect to the independent search ($p = 0$). The fact that the best performance is obtained when the imitation probability is maximum is due to the absence of local maxima in the landscape for $K = 0$. We recall that for $p = 1$ only the model string, which is likely to be represented by several copies in the group, can perform the elementary move; all other strings must imitate the model. As a result, the effective size of the search space is greatly reduced, i.e., the strings are concentrated in the vicinity of the model string which cannot accommodate many more than $M = 8$ strings without duplication of work. This is the reason we observe the degradation of the performance when the group size increases beyond its optimal value. Note that for $K = 0$ the imitative learning search always performs better than the independent search.

C. Search on a Slightly Rugged Landscape ($K = 2$)

Now we consider a somewhat more complex NK fitness landscape by setting $N = 12$ and $K = 2$. In the particular realization of the landscape studied here there are 5 maxima in total, among which 4 are local maxima. The relative fitness of those maxima (Φ/Φ_g), as well as their Hamming distances to the global maximum (d), is presented in Table I.

The results of the imitative learning search are summarized in Fig. 4 where the mean rescaled computational cost is plotted against the group size for different values of the imitation probability. The performance of the independent search ($p = 0$) is identical to that shown in Fig. 3 for $K = 0$, since that search strategy is not affected by the complexity of the landscape. The results for the cooperative search ($p > 0$), however, reveal a trade-off between the group size M and the imitation probability p. In particular, for $p > 0.8$ we observe a steep increase of the computational cost for intermediate values of M. This happens because the group can be trapped near one of the local maxima. For large groups ($M > 100$ in this case), chances are that some of the initial strings are close to the global maximum and end up attracting the rest of the group to its neighborhood, thus attenuating the effect of the local maxima. The robust finding is that for any $p > 0$ there is an optimal group size, which depends on the value of p, that minimizes the computational cost of the search.
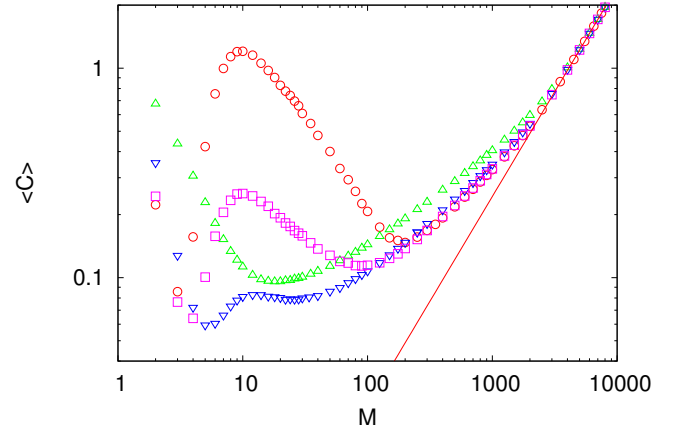

FIG. 4. (Color online) Mean rescaled computational cost $\langle C \rangle$ as function of the group size M for the imitation probability $p = 0.5$ (*triangles*), $p = 0.8$ (*inverted triangles*), $p = 0.95$ (*squares*) and $p = 0.99$ (*circles*). The solid line is the linear function $\langle C \rangle = M/2^N$. The parameters of the NK landscape are $N = 12$ and $K = 2$. The landscape exhibits 4 local maxima and a single global maximum.
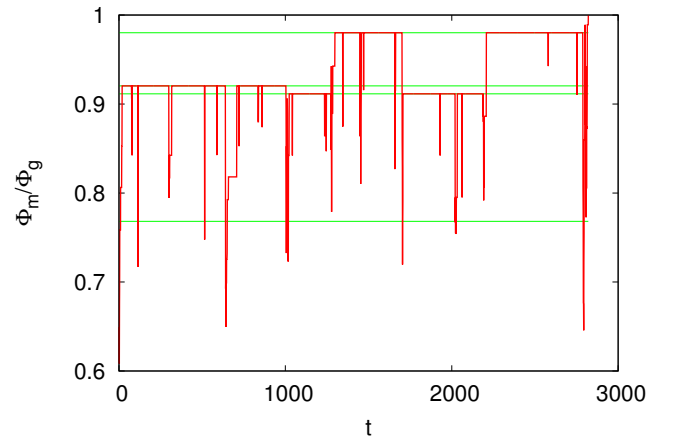

FIG. 5. (Color online) Time evolution of the relative fitness of the model string for $M = 3$ agents and probability of imitation $p = 0.99$. The thin horizontal lines indicate the relative fitness of the 4 local maxima given in Table I. The parameters of the NK landscape are $N = 12$ and $K = 2$.

A better understanding of the dynamics of the search is offered in Fig. 5 which shows the relative fitness of the model string as function of the number of trials for $M = 3$ and $p = 0.99$. The role of the local maxima as transitory attractors of the search is evident in this figure. A typical run with $M = 10$ agents, which is approximately the location of the maximum of $\langle C \rangle$ in Fig. 4, yields essentially a flat line at the highest fitness local maximum (maximum 4 in Table I) and a sudden jump to the global maximum. Because there are many copies of the model string (the mean Hamming distance between the $M = 10$ strings is less than 1), the variants produced
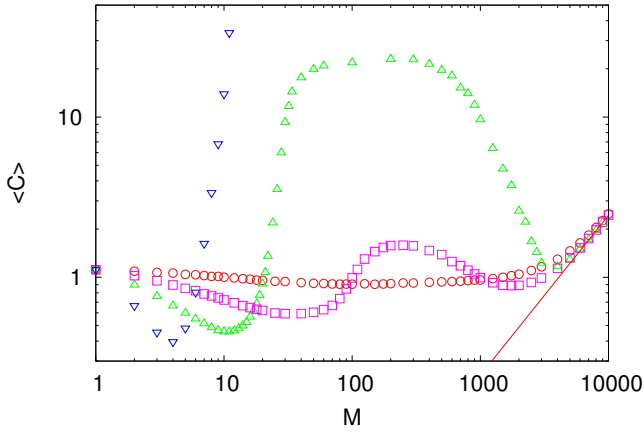
FIG. 6. (Color online) Mean rescaled computational cost $\langle C \rangle$ as function of the group size M for the imitation probability $p = 0.1$ (*circles*), $p = 0.3$ (*squares*), $p = 0.5$ (*triangles*) and $p = 0.8$ (*inverted triangles*). For $p = 0.8$ we find $\langle C \rangle > 50$ for $M > 12$. The solid line is the linear function $\langle C \rangle = M/2^N$. The parameters of the NK landscape are $N = 12$ and $K = 4$. The landscape exhibits 52 local maxima and a single global maximum.

FIG. 7. (Color online) Mean rescaled computational cost $\langle C \rangle$ as function of the probability of imitation p for the group size $M = 5$ (solid line), $M = 10$ (dashed line) and $M = 100$ (dotted line). The parameters of the NK landscape are $N = 12$ and $K = 4$. The landscape exhibits 52 local maxima and a single global maximum.

by the elementary move are attracted back to the local maximum. This is the reason why a small group can search the solution space much more efficiently than a large one in the case imitation is very frequent.

D. Search on a Rugged Landscape ($K = 4$)

The particular realization of the NK fitness landscape with $N = 12$ and $K = 4$ that we consider now has 53 maxima, including the global maximum, which poses a substantial challenge to any hill-climbing type of search strategy. As in the previous case, we observe in Fig. 6 a trade-off between M and p, but now the results reveal how imitative learning may produce disastrous performances on rugged landscapes for certain ranges of those parameters. The strategy of following or imitating a model string can backfire if the fitness landscape exhibits high fitness local maxima that are distant from the global maximum. A large group may never escape from those maxima due to the attractive effect of the clones of the model string. This is what we observed in the case of $p = 0.8$ for which we were unable to find the global maximum with groups of size $M > 12$.

We note, however, that for a fixed group size M it is always possible to tune the imitation probability p so that the imitative learning strategy performs better than (or, in a worst-case scenario, equal to) the independent search. This point is illustrated in Fig. 7 that shows the computational cost as function of p. For any fixed value of M, the computational cost exhibits a well-defined minimum that determines the value of the optimal imitation frequency. As hinted in the previous figures, this optimal
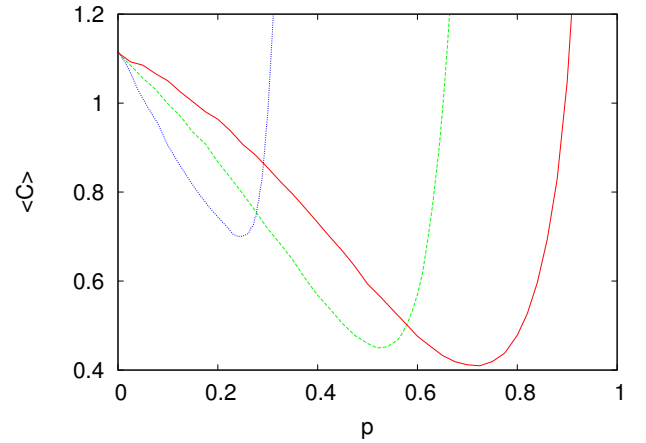
value decreases with increasing group size.

In order to verify the robustness of our findings, which were obtained for specific realizations of the NK fitness landscape, we have considered four random realizations of the landscape with $N = 12$ and $K = 4$ in addition to the realization studied above. The comparison between the mean computational costs to find the global maxima of the five realizations of the landscape is shown in Fig. 8 for the imitation probability $p = 0.5$. The results are qualitatively the same, despite the wild fluctuations of $\langle C \rangle$ in the regime where the search is trapped in the local maxima. It is reassuring to note that the initial decrease of the mean cost with increasing group size and the existence of an optimal group size that minimizes that cost, which are exhibited by all five realizations of the NK landscapes shown in Fig. 8, are robust properties of the imitative learning search.

V. DISCUSSION

In this paper we study quantitatively the problem solving performance of a group of agents capable to learn by imitation. The performance measure we consider is the computational cost to find the global maximum of three specific realizations of the NK fitness landscape. The computational cost is defined as the product of the number of agents in the group and the number of attempted trials till some agent finds the global maximum. Our main conclusion, namely, that for a fixed probability of imitation p there is a value of group size that minimizes the computational cost corroborates the findings of a similar study in which the task was to solve a particular cryptarithmetic problem [13]. Hence our conjecture that the efficacy of imitative learning could be a factor deter-
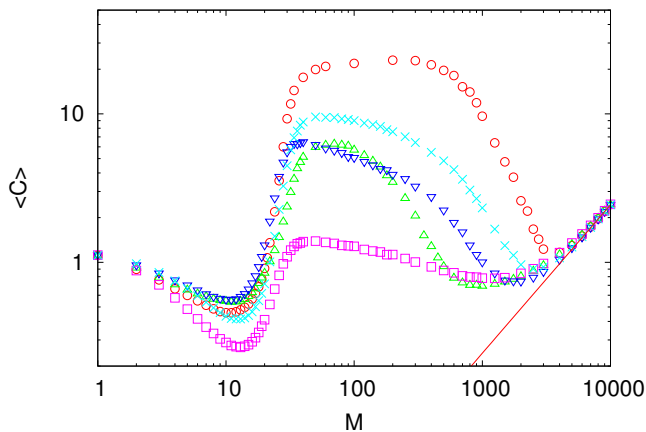
FIG. 8. (Color online) Mean rescaled computational cost $\langle C \rangle$ for the probability of imitation $p = 0.5$ as function of the group size M for five realizations (different symbols) of the NK fitness landscape with $N = 12$ and $K = 4$. The solid line is the linear function $\langle C \rangle = M/2^N$.

minant of the group size of social animals (see [1, 30] for a discussion on the standard selective pressures on group size in nature). We note that in the case the connectivity between agents is variable, i.e., each agent interacts with $L = 1, \ldots, M-1$ distinct randomly picked agent (here we have focused on the fully connected network $L = M - 1$ only) then there is an optimal connectivity value that minimizes the computational cost [13]. It would be most interesting to understand how the network topology influences the performance of the group of imitative agents and how the optimal network topology correlates with the known animal social networks [31].

The main aim of our contribution is to show that the existence of an optimal group size that maximizes performance for imitative learning [13] is insensitive to the choice of the fitness landscape, so it is likely a robust property of populations that use imitation as a cooperative strategy. Although we have focused on the effect of the parameter K, which roughly determines the rugged-

ness of the NK landscape, the parameter N (the length of the strings) also plays a relevant role on the search for the global maximum, besides the obvious role of fixing the size 2^N of the search space. (Of course, since the typical time to find the global maximum scales with 2^N, even moderate values of N, say $N = 16$, would render the simulations unfeasible.) The nontrivial role is that the value of N seems to pose an upper bound to the optimal size of the group M in the regime that imitation is very frequent (see Figs. 3, 4 and 6). This is so because in that regime the strings are concentrated in the close vicinity of the model string, which cannot accommodate more than $M = N$ different strings.

We do not purport to offer here any novel method to explore efficiently rugged landscapes, but the finding that for small group sizes imitative learning decreases considerably the computational cost of the search, even in a very rugged landscape (see data for $p = 0.8$ in Fig. 6) motivates us to address the question whether in such landscapes that strategy could achieve a better-than-random performance for all group sizes. This is achieved automatically for smooth landscapes (see Figs. 3 and 4) but not for rugged ones (see Fig. 6 and [13]). Clearly, the way to accomplish this aim is to decrease the probability of imitation p as the group size M increases, following the location of the minima shown in Fig. 7. It is interesting to note that the finding that too frequent interactions between agents may harm the performance of the group (see Fig. 7) may offer a theoretical justification for Henry Ford's factory design in which the communication between workers was minimized in order to maintain the established efficiency and maximal productivity [32].

[1] E. Wilson, *Sociobiology* (Harvard University Press, Cambridge, MA,1975)
[2] A. Whiten, D. Erdal, Phil. Trans. R. Soc. B **367**, 2119–2129 (2012)
[3] H. Bloom, *Global Brain: The Evolution of Mass Mind from the Big Bang to the 21st Century* (Wiley, New York, 2001)
[4] C.L. Nehaniv, K. Dautenhah, *Imitation and Social Learning in Robots, Humans and Animals* (Cambridge University Press, Cambridge, UK, 2007) pp. 1–18
[5] J. Kennedy, Adapt. Behav. **7**, 269–288 (1999)
[6] E. Bonabeau, M. Dorigo, G. Theraulaz, *Swarm Intelligence: From Natural to Artificial Systems* (Oxford University Press, Oxford, UK, 1999)
[7] J. Kennedy, J. Conflict Res. **42**, 56–76 (1998)
[8] J.F. Fontanari, Phys. Rev. E **82**, 056118 (2010)
[9] L. Hong, S.E. Page, J. Econ. Theory **97**, 123–163 (2001)
[10] S.E. Page, *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies* (Princeton University Press, Princeton, NJ, 2007)
[11] B.A. Huberman, Physica D **42**, 38–47 (1990)
[12] S.H. Clearwater, B.A. Huberman, T. Hogg, Science **254**, 1181–1183 (1991)
[13] J.F. Fontanari, PLoS ONE **9**, e110517 (2014)
[14] R. Axelrod, *The Evolution of Cooperation* (Basic Books, New York,1984)
[15] S. Kauffman, S. Levin, J. Theor. Biol. **128**, 11–45 (1987)

[16] S. Kauffman, E. Weinberger, J. Theor. Biol. **141**, 211–245 (1989)

[17] D. Solow, A. Burnetas, M. Tsai, N.S. Greenspan, Complex Systems **12**, 423–456 (2000)

[18] M.R. Garey, D.S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness* (Freeman, San Francisco, CA, 1979)

[19] B. Derrida, Phys. Rev. B **24**, 2613–2626 (1981)

[20] D.B. Saakian, J.F. Fontanari, Phys. Rev. E **80**, 041903 (2009)

[21] Y. Shibanai, S. Yasuno, I. Ishiguro, J. Conflict Res. **45**, 80–96 (2001)

[22] J.C. González-Avella, M.G. Cosenza, M. Eguíluz, M. San Miguel, New J. Phys. **12**, 013010 (2010)

[23] L.R. Peres, J.F. Fontanari, Europhys. Lett. **96**, 38004 (2011)

[24] L.R. Peres, J.F. Fontanari, Phys. Rev. E **86**, 031131 (2012)

[25] R. Axelrod, J. Conflict Res. **41**, 203–226 (1997)

[26] M. Perc, A. Szolnoki, BioSystems **99**, 109–125 (2010)

[27] Z. Wang, L. Wang, M. Perc, Phys. Rev. E **89**, 052813 (2014)

[28] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, *Numerical Recipes in Fortran: The Art of Scientific Computing* (Cambridge University Press, Cambridge, MA, 1992) pp. 480–486

[29] W. Feller, *An Introduction to Probability Theory and Its Applications* Vol. 1 (Wiley, New York, Third Edition, 1968) p. 220

[30] R.I.M. Dunbar, J. Human Evol. **22**, 469–493 (1992)

[31] T. Wey, D.T. Blumstein, W. Shen, F. Jordán, Anim. Behav. **75**, 333–344 (2008)

[32] S. Watts, *The People's Tycoon: Henry Ford and the American Century* (Vintage, New York, 2006)